

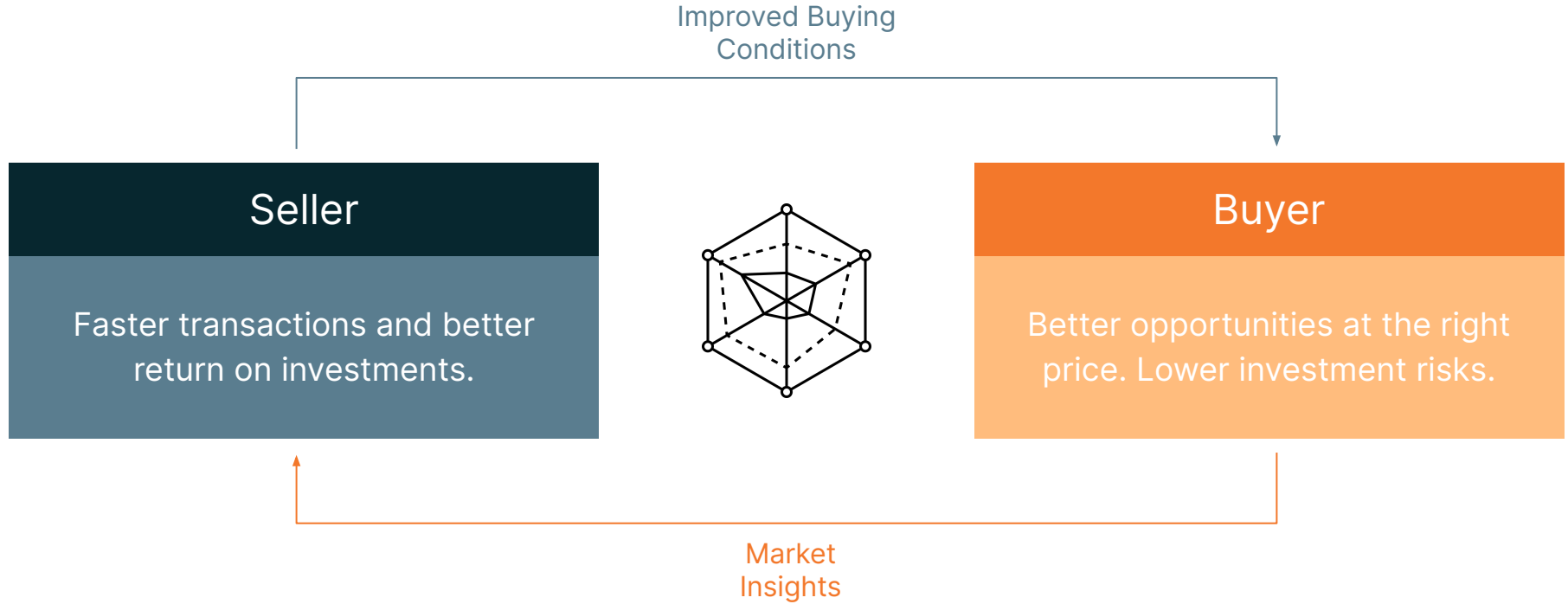
A photograph of the Seattle skyline at sunset. The Space Needle is prominent on the left, illuminated with warm lights. The city's skyscrapers are lit up, and the sky is a mix of orange, pink, and blue. Mount Rainier is visible in the background on the right.

Selling Seattle

**Fast & accurate real
estate appraisals.
Every time.**

April 2023
Ilkay Mueller, Mathilde Marie Monge

Why work with Selling Seattle?



Selling Seattle:

Leveraging data to create fast and reliable real estate appraisals. Every time.

We created a predictive model based on a linear regression to give you insights on the real estate market when you want to sell your beloved property or find your dream home.

We based our model on house sale prices for King County, which includes Seattle. It includes homes sold between May 2014 and May 2015. Therefore our target variable (which we want to predict) is the price.

Along with house prices, the data set consists of information on 18 features.



Exploratory Data Analysis (EDA):

The goal of the EDA was to gain an understanding of the real estate landscape in Seattle so that we can identify house features that affect the price variable and can be potential predictors.

Outliers

Using Outlier Boxplot, we identified that although there were outliers in the price variable, the outliers were correlated with outliers in other features (ie, sqft living, view and condition). Therefore, these outliers were good data points for our model

Correlations

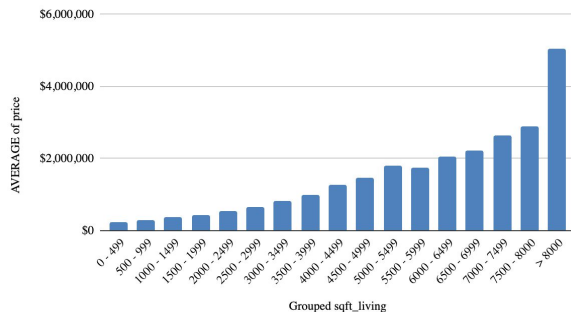
There is a high positive correlation between price and living square foot. There is a medium positive correlation between the number of bathrooms and the price. There is a low positive correlation with number of bedrooms.

Other Patterns

There is a strong positive correlation between number of bathrooms, number of bedrooms and living square footage. This can explain the correlation between price and living square footage. People want more rooms and bathrooms, which take up more space.

Data Visualisation

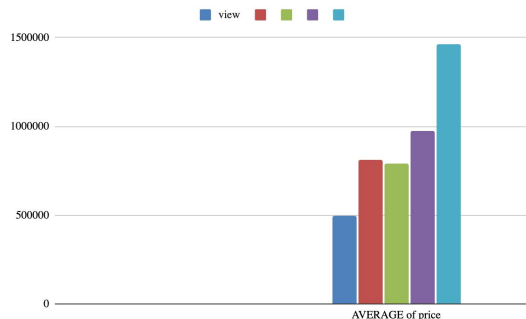
AVERAGE of price vs. Grouped sqft_living



Square Feet Living

Home prices below 500k USD have less than 2500 square feet of living area and less than 3 bathrooms.

Properties with a lower price range may also have smaller lot sizes and fewer bedrooms compared to higher-priced homes.

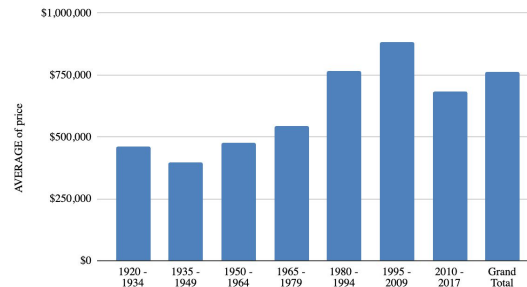


View

Location, location, location: the better the view, the higher the price.

Proximity to amenities such as shopping centers, schools, and public transportation can also significantly affect the property's value.

AVERAGE of price vs.

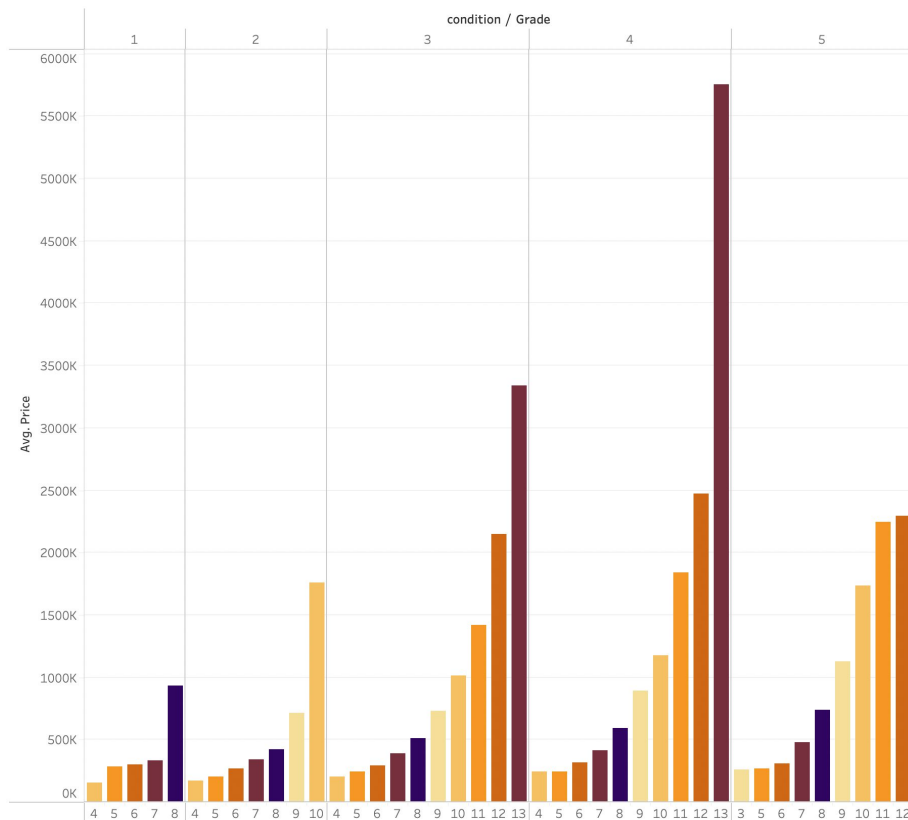


Renovations

While other factors may not be in the seller's control, some aspect of the home price can be improved: the average house price associated with **houses renovated** is up to 47% higher than the average price associated with non-renovated average house price.

Data Visualisation - Continued

Impact of Condition, Grade on Price



Condition, Grade have an impact on Price:

This indicates that the condition of a home is a significant factor in determining its value. Homes that are in better condition and have higher grades will typically command a higher price in the market.

This is because buyers are willing to pay more for homes that are move-in ready and require less maintenance or renovation. Additionally, homes in better condition may be more attractive to lenders, as they are seen as a lower risk investment.

Linear Regression:

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables by fitting a linear equation to the observed data.

Power Transformer

This is feature scaling technique that transforms the data to have a Gaussian distribution. It is particularly useful when dealing with skewed data distributions that can negatively impact the performance of certain machine learning algorithms.

R²: ~0.56 (ok)

Standard Scaler

This is a feature scaling technique that scales the data to have zero mean and unit variance. It is commonly used in machine learning to normalize the features, making them more comparable and easier to interpret.

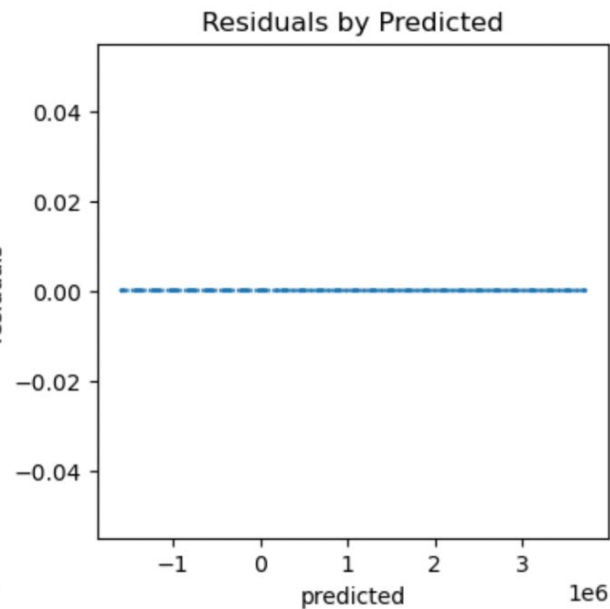
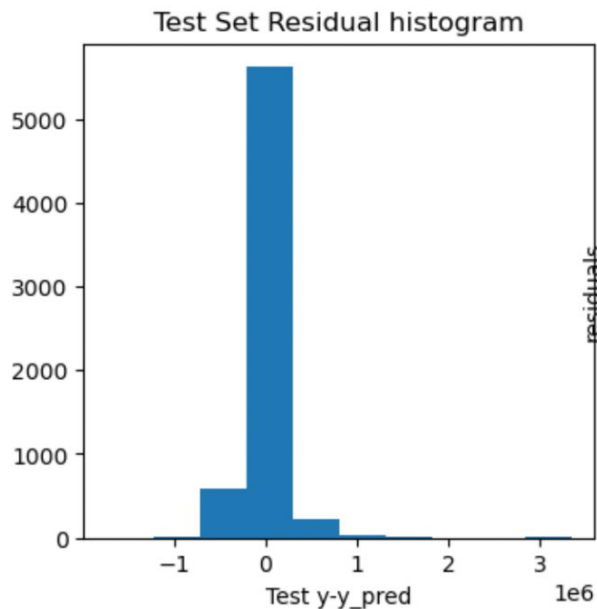
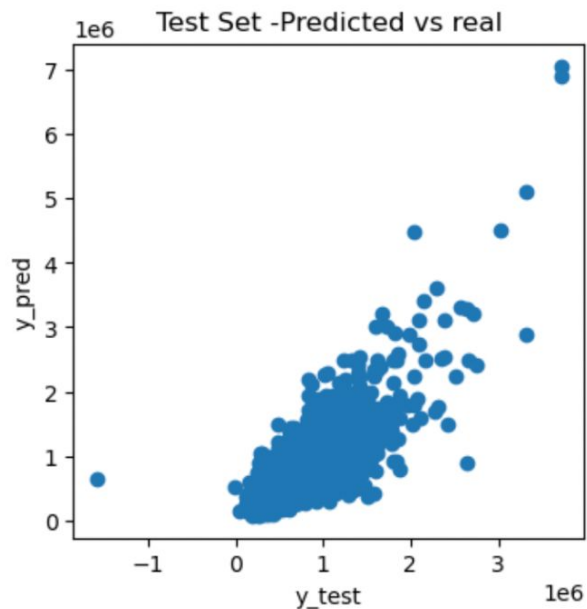
R²: ~0.66 (better)

Combination

By applying the Power Transformer first to normalize the distribution, and then using the Standard Scaler to standardize the data, we can obtain features that are normally distributed and uniformly scaled, leading to better machine learning performance.

R²: ~0.68 (best!!)

How can we tell if this is a good predictive model?



Predicting House Prices in Seattle:

Pick your dream home!



Broadway

- Condition: 4
- Grade: 7
- Living area: 1500 square feet
- 2 bedrooms
- 1 bathroom



Alaskan Way

- Condition: 2
- Grade: 9
- Living area: 4000 square feet
- 5 bedrooms
- 3 bathroom



Aurora Avenue

- Condition: 6
- Grade: 11
- Living area: 3000 square feet
- 2 bedrooms
- 2 bathroom



Selling Seattle

Ilkay Mueller
Mathilde Marie Monge